J Biomed Inform. Author manuscript; available in PMC 2013 February 1.

Published in final edited form as:
J Biomed Inform. 2012 February ; 45(1): 93–100. doi:10.1016/j.jbi.2011.09.002.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Semantic Mappings and Locality of Nursing Diagnostic Concepts in UMLS

Tae Youn Kim[a,*], **Amy Coenen**[a], and **Nicholas Hardiker**[b]

[a]College of Nursing, University of Wisconsin-Milwaukee, 1921 E. Hartford Avenue, P.O. Box 413, Milwaukee, WI 53201, USA

[b]School of Nursing & Midwifery, University of Salford, Mary Seacole Building, Greater Manchester, M6 6PU, UK

Abstract

One solution for enhancing the interoperability between nursing information systems, given the availability of multiple nursing terminologies, is to cross-map existing nursing concepts. The Unified Medical Language System (UMLS) developed and distributed by the National Library of Medicine (NLM) is a knowledge resource containing cross-mappings of various terminologies in a unified framework. While the knowledge resource has been available for the last two decades, little research on the representation of nursing terminologies in UMLS has been conducted. As a first step, UMLS semantic mappings and concept locality were examined for nursing diagnostic concepts or problems selected from three terminologies (i.e., CCC, ICNP, and NANDA-I) along with corresponding SNOMED CT concepts. The evaluation of UMLS semantic mappings was conducted by measuring the proportion of concordance between UMLS and human expert mappings. The semantic locality of nursing diagnostic concepts was assessed by examining the associations of select concepts and the placement of the nursing concepts on the Semantic Network and Group. The study found that the UMLS mappings of CCC and NANDA-I concepts to SNOMED CT were highly concordant to expert mappings. The level of concordance in mappings of ICNP to SNOMED CT, CCC and NANDA-I within UMLS was relatively low, indicating the need for further research and development. Likewise, the semantic locality of ICNP concepts could be further improved. Various stakeholders need to collaborate to enhance the NLM knowledge resource and the interoperability of nursing data within the discipline as well as across health-related disciplines.

Keywords

Unified Medical Language System; Standardized Nursing Terminology; Semantic Mapping; Nursing Knowledge Representation; Health Informatics

*Corresponding Author: 1921 E. Hartford Avenue, P.O. Box 413, College of Nursing, University of Wisconsin-Milwaukee, Milwaukee, WI 53201-0413, USA, <kimty@uwm.edu>, Tel: +1 414 229 3986, Fax: +1 414 229 6474.

1. Introduction

Electronic health records (EHRs) are considered a means to support decision-making of healthcare providers, administrators, as well as consumers, thereby ensuring high quality of care, safe practice, and cost reduction [1]. According to the Health Information Technology for Economic and Clinical Health (HITECH) Act (which was part of the American Recovery and Reinvestment Act of 2009), specifications on meaningful use of EHRs have been established in the U.S. [2]. The specifications, for example, include the use of computerized provider order entry (CPOE) for medication orders, maintenance of problem and medication lists, and electronic exchange of clinical information [3]. The Centers for Medicare & Medicaid Services (CMS) subsequently announced an incentive program for certified EHRs meeting standards and criteria issued by the Office of the National Coordinator for Health Information Technology [3]. It would appear that standards that facilitate data capture and sharing are a foundation in promoting the meaningful use of patient care data.

Nursing forms the largest group of healthcare providers in the U.S. and plays a key role in ensuring the quality of care in collaboration with other health care providers and healthcare consumers [4]. It is important to maintain accurate and consistent descriptions of patient care using standardized nursing terminologies. There are two minimum data sets and 10 terminologies or classifications recognized by the American Nurses Association as standards that apply to nursing practice [5]. Of the 10 terminologies or classifications, ABC Codes, Logical Observation Identifiers Names and Codes (LOINC), and Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) are considered multidisciplinary terminologies while the other seven are recognized as interface nursing terminologies. These seven are Clinical Care Classification (CCC; formerly known as Home Health Care Classification), International Classification for Nursing Practice (ICNP), NANDA-International (NANDA-I), Nursing Intervention Classification (NIC), Nursing Outcome Classification (NOC), Omaha System, and Perioperative Nursing Data Set (PNDS) [5–12]. These terminologies have been widely adopted across settings nationally and internationally in both paper-based and electronic format [6–12].

Efforts to integrate these nursing terminologies with SNOMED CT have been made; the integration of ICNP is currently in progress in collaboration with the International Health Terminology Standards Development Organisation (IHTSDO) [13]. Furthermore, all of these terminologies have been added to the Unified Medical Language System (UMLS) designed by the U.S. National Library of Medicine (NLM) [14]. As a knowledge resource, the UMLS provides a unified platform to harmonize all source vocabularies submitted, while maintaining the original structure of the sources [15]. Regardless of a source of a term, if the semantics of a new term are identical to an existing term in UMLS, the new term is assigned to the same concept unique identifier (CUI), indicating the two terms are equivalent in meaning [16–17].

Although each of the seven nursing terminologies was designed for a particular purpose, there is significant overlap; identical and synonymous concepts need to be harmonized when integrated with SNOMED CT and UMLS. We hypothesized that if a semantically-equivalent concept appears in multiple nursing terminologies, the concept should be assigned the same concept identifier in SNOMED CT and in UMLS. This would ensure that equivalent nursing concepts are exchanged in a consistent manner across different systems. Initially, we designed this study to examine the extent to which select nursing concepts were accurately represented in the UMLS. Nursing diagnostic concepts or problems selected from CCC, ICNP and NANDA-I were of particular interest in this study.

To our knowledge, there is little research on the representation of nursing domain knowledge in UMLS. Thus, this study aimed to (a) determine the concordance of semantic mappings of the selected diagnostic concepts in the unified framework; (b) exploit synonymous relationships of diagnostic concepts across the source terminologies; and (c) explore the placement of the selected nursing concepts on the UMLS semantic network. It is anticipated that this study will inform nursing terminology researchers and users on the utility of the knowledge source and the areas of further research and development.

2. Background

2.1. Nursing Diagnostic Concepts or Problems

Of the seven nursing terminologies above, four terminologies – CCC, ICNP, NANDA-I and PNDS – include nursing diagnostic concepts or problems representing a clinical judgment on the healthcare needs of individual, family and/or community given his/her illness [8]. Such clinical judgment becomes a basis for deriving nursing care plans with ongoing assessments, outcome evaluations, and interventions targeted to a given problem [8].

The nursing terminologies have much in common and yet there are differences. For example, the diagnostic concepts of PNDS are identical to the NANDA-I [12] and certain CCC nursing diagnostic concepts were adapted from NANDA-I [6]. There exist some differences between the terminologies, however. For instance, in terms of internal organization, NANDA-I (2009–2011) comprises 201 nursing diagnostic concepts, which are classified into 47 classes and, 13 domains [8]. CCC Version2 is comprised of 182 nursing diagnoses, 59 major and 123 subcategory concepts. These 182 CCC diagnoses are grouped into 21 care components which are then further clustered into functional, health behavioral, physiological and psychological patterns [6]. To illustrate the differences between the terminologies, *Activity Intolerance* is a nursing diagnostic concept within both NANDA-I and CCC, representing a patient's lack of ability or energy to endure or complete activities in daily living. In NANDA-I, this concept is classified as a 'Cardiovascular/Pulmonary Responses' within the 'Activity/ Exercise' domain [8]; in CCC the concept falls within the 'Activity Component' [6].

The way in which a concept is defined and classified in ICNP is different yet again from the conceptual schemas employed in the other nursing terminologies. The ICNP is designed using Web Ontology Language (OWL) in the Protégé ontology development environment. Each concept in ICNP is formally defined using description logic and is classified automatically by a reasoner according to its formal properties [18]. The 2011 release of ICNP includes 669 pre-coordinated diagnostic and outcomes concepts, which are in turn composed from 2,136 primitive concepts. Using the same *Activity Intolerance* example, the pre-coordinated diagnostic concept Activity Intolerance (10000431) is composed with one primitive focus concept Activity Intolerance (10000408) and two primitive judgment concepts – Actual (10000420) and Impaired (10012938). Accordingly, the pre-coordinated ICNP concept represents the same diagnosis as CCC and NANDA-I. In ICNP, however, there are no organizing categories such as the classes and domains of NANDA-I or Care Components of CCC. ICNP concepts are classified instead according to the related formal properties.

Even though there may be differences between the terminologies, it should be noted that all of these terminologies conform to the International Organization for Standardization (ISO) technical standard18104: integration of a reference terminology model for nursing (18104:2003) [19] which facilitates consistent description of concepts being represented within and across terminologies or electronic health record systems. Moreover, the ISO standard has promoted, in part, the development of a cross-mapping methodology across

NIH-PA Author Manuscript

NIH-PA Author Manuscript

nursing terminologies by decomposing a diagnostic concept into required components such as focus and judgment [20– 21].

2.2. The Unified Medial Language System (UMLS)

Since 1990, the UMLS has served as a knowledge resource in the biomedical and health informatics communities. The UMLS contains a vast number of concepts and relations that exist within and across terminologies [16–17]. Source vocabularies submitted have diverse structures ranging from a plain taxonomy with hundreds of concepts to a logic-based terminology system with hundreds of thousands of concepts. The UMLS merges such diverse terminologies in a unified framework while maintaining the original structural variations [15].

Two components of UMLS include the Metathesaurus and the Semantic Network determined by semantic types and their relations [16]. The Metathesaurus currently contains more than two millions concepts submitted from more than 130 source vocabularies such as MeSH, ICD-9CM, ICD-10, ICF, SNOMED CT, LOINC, CCC, NANDA-I, and ICNP [14]. A set of automated and semi-automated techniques are employed to integrate a source vocabulary into the unified system, along with human review [17]. In this integration process, each individual concept is assigned to a concept unique identifier (CUI). If a concept from a source vocabulary exists in another source vocabulary, the concepts with the same meaning are considered identical and assigned to the same CUI. For example, a concept *Acute Pain* appears in multiple terminologies and is linked through the same CUI (C0184567) in the UMLS framework; similarly with *Chronic Pain* (C0150055).

It is thus possible to examine cross-mappings of a concept across various source terminologies using CUIs. Further, the semantic locality of concepts can be exploited based on concept associations and their placement in the Semantic Network [16–17]. Relationships among concepts within a source vocabulary and across terminologies can be viewed through CUIs in the MRREL table containing related concepts within UMLS [22]. For instance, an ICNP concept ‘C0015967 Hyperthermia’ is related to a SNOMED CT concept ‘C1704628 Body temperature above reference range’. Although the two concepts were assigned to two different CUIs, the two concepts were defined as synonymous in MRREL. In addition, parent-child and siblings relationships of concepts can be found in MRREL. For example, according to MRREL, while ‘C0184567 Acute Pain’ is a type of *Pain* in ICNP, the same concept with the same CUI, is a type of *Comfort Alteration* in CCC and NANDA-I.

Given the complexity of the UMLS, it is challenging to exploit the semantic location of existing concepts. The Semantic Network is an upper-level ontology designed to reduce the complexity of concepts and provide a high-level conceptual framework of such concepts [16–17]. Each concept can be understood in a simple, abstract level by examining one or more semantic types assigned to the concept. The 2010AB release of UMLS contains 133 semantic types and 54 semantic relations connecting the semantic types. Two root semantic types are ‘Entity’ and ‘Event’. Each root concept has a tree structure in which a child semantic type is attached to a parent semantic type through an *isa* link [22]. For example, the concept Acute Pain is assigned to a semantic type ‘sign or symptom’ (ST1) which is the *evaluation of* (rel) other semantic types such as ‘physiologic function’ (ST2) or ‘pathologic function’ (ST2) in the Semantic Network. In this example, one semantic type (ST1) has a relationship with another semantic type (ST2), which is denoted as triplets (ST1, *rel*, ST2) [22].

Both the Metathesaurus and the Semantic Network have been utilized in information retrieval and natural language processing of biomedical literature and electronic health records containing diverse expressions of a concept [23–26]. Clustering concepts with semantic similarity has been shown to expand the search space and yield a better outcome as compared to the results of a simple lexical search. The quality of information within both the Metathesaurus and the Semantic Network will obviously impact on the performance of any dependent information retrieval system [27]. In this study, we examined how nursing domain knowledge has been represented in UMLS with a particular interest in semantic mappings and locality of nursing diagnostic concepts or problems.

3. Methods

According to the purpose of the study, nursing diagnostic concepts were selected from CCC, ICNP, NANDA-I, and SNOMED CT to examine the representation of those concepts with semantic equivalence or similarity in UMLS. Figure 1 depicts the sources of data and the number of concepts examined for our analysis. All CCC and NANDA-I diagnostic concepts available in UMLS were included in this analysis while only ICNP and SNOMED CT concepts that were mapped to CCC and NANDA-I were selected. Along with the UMLS 2010AB release, the cross-mappings performed by human experts were utilized as a reference standard for our analysis. In other words, expert cross-mappings between CCC and ICNP, NANDA-I and ICNP, CCC and SNOMED CT, NANDA-I and SNOMED CT, as well as ICNP and SNOMED CT became the sources for evaluation of UMLS semantic cross-mappings among the nursing diagnostic concepts selected.

It should be noted that mappings of CCC and NANDA-I to ICNP were completed by terminology experts as a part of previous studies [20,28]. Mappings of CCC and NANDA-I to SNOMED CT were conducted by terminology experts of CCC and NANDA-I in collaboration with the College of American Pathologists (CAP)'s SNOMED CT team [6,8,29]. The cross-mappings of CCC and NANDA-I concepts with SNOMED CT are publicly available for use [29]. The authors extracted the mapping information from the ClueData (SNOMED CT International Release 2010–07–31) via the CliniClue browser [30]. Mappings between ICNP and SNOMED CT concepts included in this study were completed by the authors as part of a larger IHTSDO project [13]. The ICNP-SNOMED CT mapping methodology and results will be reported elsewhere. Due to the fact that both CCC and NANDA-I were mapped to ICNP and SNOMED CT, it was considered that there might be high-level of agreement between cross-mappings performed by human experts and cross-mappings available in UMLS. Based on this assumption, we identified five major hypotheses for which the study methods were developed as described below.

3.1. Hypothesis 1

A nursing diagnostic concept is unique in each source terminology (i.e., CCC, ICNP, and NANDA-I) so that there will be no duplicate CUIs within a source terminology in UMLS. This hypothesis was tested by examining if there is a CUI assigned to more than one concept in each source terminology.

3.2. Hypothesis 2

Concepts that are semantically equivalent between CCC and SNOMED CT, and between NANDA-I and SNOMED CT will appear to be identical in UMLS. To test this hypothesis, first, we identified CUIs assigned to CCC and NANDA-I concepts and then compared them to the CUIs of SNOMED CT concepts that were cross-mapped by human experts. We assumed that there might be high-level agreement of CUIs between CCC and SNOMED CT and between NANDA-I and SNOMED CT. The proportion of concordance between UMLS and expert cross-mappings was then calculated by dividing the observed number of identical concepts in UMLS by the number of identical concepts determined in human cross-mappings (1). We further assumed that if there is high-level agreement between the two cross-mappings then CUIs assigned to both terminologies were accurate for this study.

$$Concordance\ of\ UMLS\ Mappings\ (\%)=\frac{Number\ of\ Identical\ Concepts\ in\ UMLS}{Number\ of\ Identical\ Concepts\ in\ Expert\ Mappings}\times 100 \quad (1)$$

3. 3. Hypothesis 3

Similarly, we hypothesized that concepts that are semantically equivalent between CCC and ICNP and between NANDA-I and ICNP would appear to be identical in UMLS. We used the same method described in Hypothesis 2 to test this hypothesis. The proportion of concordance between UMLS and expert cross-mappings was calculated using the same formula above (1). It was also expected that there would be high-level of agreement between the two cross-mappings and thus, possibly driving a mapping table between selected ICNP concepts and SNOMED CT. Note that cross-mappings of ICNP diagnostic concepts to concepts within NANDA-I, CCC and SNOMED CT are not yet publicly available.

3. 4. Hypothesis 4

If concepts with semantic equivalence are not linked through CUIs in UMLS, such information might be represented through synonymous (SY), broader (RB), narrower (RN), or possibly synonymous (RQ) relationships in the UMLS table (MRREL). This hypothesis was tested by exploiting the distribution of concepts that were related to each other through the four types of relationships. We also compared these findings to the cross-mappings table created by human experts for a reference standard.

3.5. Hypothesis 5

Nursing diagnostic concepts selected from CCC, ICNP, and NANDA-I will be mapped to the semantic type ‘Finding’ or similar categories as a nursing diagnostic concept is a clinical judgment on the care needs of individual, family or community. In addition, it was assumed that if one nursing diagnostic concept is semantically equivalent to another diagnostic concept, both concepts should be assigned to the same semantic type. We evaluated the distribution of semantic types assigned to nursing diagnostic concepts selected from the three source terminologies.

4. Results

This section summarizes our findings associated with the five hypotheses described above. Findings related to Hypotheses 1 to 3 summarize the semantic mappings of nursing diagnostic concepts in UMLS whereas findings related to Hypotheses 4 and 5 present the semantic locality of the concepts. We examined a total of 857 concepts involving 182 concepts from CCC, 167 concepts from NANDA-I, 239 concepts from ICNP, and 269 concepts from SNOMED CT (Figure1).

4.1. Hypothesis 1

Since each source terminology contained a set of unique concepts, we expected that each concept within a source would be assigned to a unique CUI in UMLS. As shown in Table 1, however, there were duplicate CUIs in each source terminology; 2.4% of CCC concepts, 1.7% of ICNP concepts, and 1.2% of NANDA-I concepts. The pairs of concepts with the same CUI indicate that those concepts were either erroneously treated as concepts with the same meaning in UMLS or were not unique as assumed by the terminology developers.

4.2. Hypothesis 2

According to the expert mappings, there were 100% of semantic matches between CCC and SNOMED CT and between NANDA-I and SNOMED CT. Table 2, however, shows some discrepancies between UMLS and expert cross-mappings, particularly for ICNP.

According to the mappings via CUIs, 20 concepts (11%) from CCC were not semantically comparable to SNOMED CT concepts in UMLS. Of these 20 non-matched concepts, seven concepts were associated with a *Knowledge Deficit* of a client. For example, a CCC concept 'D08.5 Knowledge Deficit of Medication Regimen' was mapped to '129866007 Deficient knowledge of medication regimen (finding)' in SNOMED CT but the two concepts were assigned to two different CUIs - C0184529 and C1268751 respectively. On the other hand, four NANDA-I concepts (2.4%) were not mapped to SNOMED CT through CUIs in UMLS (Table 2). For instance, a NANDA-I concept '00088 Impaired walking' was mapped to '228158008 Walking disability (finding)' in SNOMED CT but the two concepts were assigned to two different CUIs - C0311394 and C0556291 respectively.

Overall, the level of concordance between UMLS and expert mappings with regard to CCC and NANDA-I was high ranging from 89% to 98% so that the UMLS could be considered a reliable knowledge source from the nursing domain perspective. Interestingly, only 28% of ICNP concepts, however, were mapped to SNOMED CT in UMLS through CUIs, which was substantially different from the expert mappings presenting 93% of matches between the two terminologies in the subset used for this study. It is not clear why the level of concordance between the UMLS and human expert mappings is so low (30%) for this particular terminology.

4.3. Hypothesis 3

According to the mappings performed by human experts, 95% of CCC concepts and 99% of NANDA-I concepts were matched to ICNP concepts. Table 3 presents the results of comparison between UMLS and expert mappings conducted through ICNP. When examined the level of concordance between the two mappings, a similar phenomenon observed in testing the second hypothesis was detected. That is, only 22.5% of CCC concepts and 40% of NANDA-I concepts were cross-mapped to ICNP in UMLS (Table 3). Examples of agreement and disagreement between the two mappings are presented in Table 4.

4.4. Hypothesis 4

The hierarchical relationships among the concepts in each terminology were accurately captured in MRREL as they are designed by the terminology developers. These relationships included parent-child and sibling relationships between concepts. When the four types of synonymous relationships (SY, RB, RN, and RQ) in UMLS were examined, there were a relatively small number of synonymous concepts across the terminologies as shown in Table 5. Of 182 CCC concepts, 36% of the concepts presented synonymous relationships with ICNP and 21% with SNOMED CT active concepts, leaving 55% of CCC concepts not associated with nursing diagnostic concepts from ICNP, NANDA-I, or SNOMED CT. When the synonymous concepts were examined, 13 ICNP concepts were erroneously associated with CCC and five CCC concepts did not correspond with the SNOMED CT concepts identified by human experts. For example, a CCC concept 'K25.3 Hypothermia' had a synonymous relationship with an ICNP primitive concept '10009547 Hypothermia' in

MRREL. According to the ICNP structure, however, the CCC concept should be mapped to a pre-coordinated diagnostic concept ‘10000761 Hypothermia’ instead of the ICNP primitive concept.

Similarly, 49% of NANDA-I concepts had synonymous relationships with ICNP, while 20% of NANDA concepts had relationships with SNOMED CT, leaving 47% unassociated with other concepts from CCC, ICNP or SNOMED CT. When those synonymous concepts were examined, six ICNP concepts were incorrectly associated with NANDA-I and one NANDA concept did not correspond with the SNOMED concept identified by human experts. For example, a NANDA-I concept ‘00101 Adult Failure to Thrive’ was semantically equivalent to a SNOMED CT concept ‘129588001 Adult failure to thrive syndrome (disorder)’ through a CUI but also had a ‘narrower than’ relationship with the SNOMED CT concept in MRREL.

Although CCC and NANDA-I had synonymous relationships with ICNP, no ICNP concepts were identified as synonyms with CCC and NANDA-I in UMLS. Only 8% of ICNP concepts were determined as synonyms with SNOMED CT active concepts as shown in Table 5. Unexpectedly, there were 75 ICNP concepts (31%) that were related to ICNP itself. For example, a concept ‘10000477 Anxiety’ had a synonymous relationship with both ‘10000477 Anxiety’ (i.e. the same concept) and ‘10002429 Anxiety’. It appeared that the ICNP nursing diagnostic preferred term (10000477 Anxiety) and its formal OWL knowledge name (10000477 ActualNegativeAnxiety) were treated within UMLS as a synonym. It also appeared that the primitive concept (10002429 Anxiety) used to compose the nursing diagnostic concept (10000477 ActualNegativeAnxiety) in ICNP was treated within UMLS as a synonym.

4.5. Hypothesis 5

All the nursing diagnostic concepts examined in this study were allocated to only one semantic type. As summarized in Table 6, most concepts (ranging from 66.5% to 81.6%) were placed on a semantic type ‘Finding’ Some nursing diagnostic concepts (ranging from 3.3% to 8.2%) were classified as ‘Sign or Symptom,’ which is a child semantic type of ‘Finding’ as depicted in Figure 2. The remaining concepts fell into the ‘Event’ root class. When the semantic types were further clustered using 15 semantic groups [31], the majority of concepts (≥ 95%) in each terminology could be clustered into ‘Disorders’ group as presented in Table 6. This result is consistent with the representation of CCC and NANDA-I concepts in SNOMED CT. That is, 99.5% of nursing diagnostic concepts fell into either the ‘Finding’ or ‘Disorder’ category in SNOMED CT except one CCC concept ‘Dying process (observable entity)’.

5. Discussion

This study examined how nursing domain knowledge is represented in UMLS with a particular focus on nursing diagnostic concepts or problems selected from CCC, ICNP, and NANDA-I. Each UMLS concept is defined by its relationships to other concepts and its placement in the Semantic Network [16]. We focused on semantic mappings and locality of the nursing diagnostic concepts in UMLS and tested five hypotheses. CCC and NANDA-I were previously cross-mapped by human experts into both ICNP and SNOMED CT. These cross mappings formed the basis of evaluation in this study. The remainder of this paper will present key findings along with implications and future research.

5.1. Existence of duplicate concepts

There were a relatively small number of concepts with the same CUIs within each nursing terminology. The two nursing concepts - *Risk for Injury* and *Risk for Trauma* - were not considered independent concepts in UMLS; rather they were considered identical concepts as they are in SNOMED CT. *Trauma* is commonly defined as “a deeply distressing or disturbing experience” followed by a “stressful event or physical *injury*” [32]. Within medicine, *trauma* and *injury* are generally considered to be equivalent. Within nursing, these two concepts are generally considered to be similar but not necessarily equivalent. Even nursing terminology experts, however, do not agree on the relationship between the two concepts. For example, the two concepts have a parent-child relationship in both CCC and ICNP. The concept *injury* is a type of *trauma* in ICNP while *trauma* is a type of *injury* in CCC. Further clarification is necessary for the clinical use of the two concepts in nursing.

5.2. Discrepancy in semantic mappings

The testing of Hypotheses 2 and 3 suggested that the way in which ICNP was integrated into UMLS might be dissimilar with the way in which CCC and NANDA-I were added. When examining the semantic mappings of nursing diagnostic concepts through CUIs in UMLS, 89% of CCC and 98% of NANDA-I concepts were cross-mapped to SNOMED CT while only 28% of ICNP concepts were mapped to SNOMED CT. In addition, only 23% of CCC and 40% of NANDA-I concepts were cross-mapped via UMLS to ICNP. This was substantially different from expert mappings which presented approximately 97% of matches on average. These findings were contradictory to our assumption that there would be a high level of concordance in CUIs of the nursing diagnostic concepts with semantic equivalence in UMLS. In fact, since CCC and NANDA-I were integrated into UMLS prior to the addition of ICNP to the UMLS, we assumed that CUIs assigned to CCC and NANDA-I would be the same as CUIs assigned to corresponding ICNP concepts. This was not the case as demonstrated in this analysis.

It is also interesting to note that expert mappings indicated that 97 concepts were semantically identical across the three nursing terminologies. Of these, 76 concepts (78.4%) were cross-mapped in UMLS through SNOMED CT concepts. In contrast, only 38% of the 97 concepts were cross-mapped in UMLS through ICNP concepts. All the findings revealed that there were large discrepancies between CUIs assigned to CCC and NANDA-I with CUIs allocated to ICNP. These discrepancies also explain the small number of matches across all three nursing terminologies in UMLS as compared to the expert mappings.

Upon further investigation of the types of discrepancy between UMLS semantic mappings and expert mappings, we found some issues that might have impacted the integration of the ICNP within the UMLS. For example, in nursing, a term ‘impaired’ is seen as equivalent in meaning to terms such as ‘ineffective’ and ‘alteration.’ These terms are used interchangeably in the three nursing terminologies studied. Likewise, a term ‘improved’ is seen as equivalent to the term ‘effective.’ Accordingly, it was thought that use of such qualifiers might have been the cause of partial matches or even no matches in UMLS between semantically equivalent concepts that use different qualifiers. In addition, the structural complexity of the ICNP polyhierarchy may have hindered the accurate representation of ICNP in the more uniform framework of UMLS. Regardless of the cause, further investigation is needed to identify why ICNP nursing diagnostic concepts that are semantically equivalent to CCC, NANDA, and SNOMED CT concepts were not detected when integrated into UMLS.

5.3. Issues related to semantic locality of nursing problems

The semantic locality of a concept is determined in part by concept relationships and placement of the concept on the Semantic Network [16]. With regard to semantic relationships, the discrepancy in integration into UMLS was observed once more when we exploited the relationship table to inspect whether the nursing diagnostic concepts were considered synonyms in UMLS. In the relationship table, most ICNP concepts (92%) were not associated with CCC, NANDA-I, or SNOMED CT. This was not the case for the other terminologies (i.e., CCC and NANDA-I). To complicate the situation further, 31% of ICNP concepts were associated with themselves. This is further evidence of issues with integration of ICNP in UMLS.

Our understanding is that the UMLS uses lexical resemblance techniques and human reviews based on rules and thus, a synonym in a source vocabulary is not necessarily a synonym in the UMLS [22]. McCray and Nelson [16] noted that synonyms are loosely defined in the UMLS for practical purposes, meaning related terms are considered synonyms if they appear closely in a given context. That is, semantically equivalent concepts in a given domain are not necessarily regarded as synonyms by convention. This rationale may go some way toward explaining the discrepancies in our findings.

The UMLS semantic network provides an abstract view of what may sometimes be large and complex vocabularies. When the semantic types of the nursing diagnostic concepts were further clustered into 15 semantic groups, more than 95% of concepts from each terminology were grouped into ‘Disorders,’ including the six semantic types – Finding, Sign or Symptom, Pathologic Function, Disease or Syndrome, Mental or Behavioral Dysfunction, and Injury or Poisoning (Table 6). The remaining five percent of concepts raise questions for further examination. For example, an ICNP concept ‘10000703 Fear (C0015726)’ is mapped to the CCC concept ‘P41 Fear (C1961839)’ and the NANDA-I concept ‘00148 Fear (C1961839)’ according to expert mappings. These concepts, however, were assigned to different semantic types; ‘Mental Process (event)’ for the ICNP concept and ‘Finding (entity)’ for CCC and NANDA-I concepts. The ‘Mental Process’ semantic type falls into the ‘Physiology’ semantic group rather than ‘Disorders’ group. Although the different CUIs explain the discrepancy in part, the assignment of semantic types also needs to be examined more closely.

5.4. Implications of the study

The UMLS Metathesaurus has been utilized to enhance information retrieval, for example, of scientific literature and electronic health records containing diverse expressions for similar concepts [23–26]. Inaccurate assignment of concept identifier and semantic type may affect the performance of information retrieval and natural language processing necessary for semantic processing of textual documents using UMLS [33–34]. This implies that auditing the representation of a specific domain concept within the unified framework is at least as important as adding a given terminology to the UMLS. Previous studies have discussed how the UMLS might be audited using multiple tools and techniques such as object-oriented modeling, metaschema, or neighboring method [35–38]. Little research, however, has been conducted to audit the UMLS for nursing domain knowledge.

It is hoped that our findings will inform terminology developers so that they might make enhancements. For instance, recently, an effort has been made by the IHTSDO Nursing Working Group Problem List Project Team in collaboration with the NLM to identify a set of nursing problems that could be utilized across nursing specialties and settings [39]. The project team consisting of nurse researchers, terminology developers, and NLM staff members developed a set of nursing problems based on the nursing diagnostic concepts mapped to SNOMED CT through CUIs in UMLS. The recent release of the list, however, suggests that the mappings of nursing problems to SNOMED CT are not comprehensive as demonstrated to some extent in this study. In light of this, the UMLS semantic mappings of nursing diagnostic concepts need to be verified by stakeholders in order to enhance the nursing problem list. The nursing profession should continue to participate in this endeavor to ensure that the domain concepts are represented in the approved manner within the UMLS construct.

6. Conclusion

There exist overlapping concepts across the nursing terminology systems, which require on-going collaborations among the terminology developers to augment the interoperability of nursing data. This investigation provides an insight into the extent to which the nursing diagnostic concepts selected from nursing terminologies are represented in UMLS and the consistency of that representation. It has shown that there are areas for improvement of UMLS that are necessary to enhance cross-mapping between nursing problems and to further the applicability of any derived nursing problem list. We recommend that the wider nursing community be proactive in collaborating with the National Library of Medicine in order to represent accurately nursing domain concepts within UMLS.

Acknowledgments

This study was supported in part by the U.S. National Institute for Nursing Research, National Institutes of Health (Grant 1R03NR012802; TK, AC) and the International Council of Nurses (ICN; TK, AC, NH).

References

1. U.S. Department of Health & Human Services. Electronic Health Records and Meaningful Use. 2011. Available from: http://healthit.hhs.gov/portal/server.pt?open=512&objID=2996&mode=2
2. Blumenthal D. Launching HITECH. N Engl J Med. 2010; 362:382–385. [PubMed: 20042745]
3. Centers for Medicare & Medicaid Services (CMS). Eligible Professional Meaningful Use Core Measures. 2010. Available from: https://www.cms.gov/EHRIncentivePrograms/Downloads/EP-MU-TOC-Core-and-MenuSet-Objectives.pdf
4. Dunton, N.; Montalvo, I. Transforming Nursing Data into Quality Care: Profiles of Quality Improvement in U.S. Healthcare Facilities. Silver Spring (MD): American Nurses Association Inc; 2007.
5. American Nurses Association. ANA Recognized Terminologies and Data Element Sets. 2006. Available from: http://www.nursingworld.org/npii/terminologies.htm
6. Saba, V. Clinical Care Classification (CCC) System Manual: A Guide to Nursing Documentation. New York City (NY): Springer Publishing Company; 2007.
7. International Council of Nurses. International Classification for Nursing Practice Version 2. Geneva, Switzerland: ICN; 2009.
8. Herdman, TH., editor. North American Nursing Diagnosis Association-International (NANDA-I) Nursing Diagnoses 2009–20011: Definitions and Classification. West Sussex, UK: Wiley-Blackwell; 2009.
9. Bulechek, GM.; Butcher, HK.; Dochterman, JM., editors. Nursing Interventions Classification (NOC). 5th ed.. St. Louis (MO): Mosby Inc; 2008.
10. Moorhead, S.; Johnson, MM.; Maas, M., editors. Nursing Outcomes Classification (NOC). 4th ed.. St. Louis (MO): Mosby Inc; 2008.
11. Martin, KS. The Omaha System: A Key to Practice, Documentation, and Information Management. Reprinted 2nd ed.. Omaha (NE): Health Connections Press; 2005.
12. Association of Operating Room Nurses, PNDS - Perioperative Nursing Data Set. 3rd ed.. Denver (CO): AORN Inc; 2011.

13. International Health Terminology Standards Development Organisation. ICN and IHTSDO Team-up to Ensure a Common Health Terminology. 2010. Available from: http://www.ihtsdo.org/fileadmin/user_upload/Docs_01/Press_Releases/ICN_and_IHTSDO_Announcement_March_2010.pdf
14. National Library of Medicine, UMLS Source Release Documentation. 2011. Available from: http://www.nlm.nih.gov/research/umls/sourcereleasedocs/
15. National Library of Medicine. UMLS Reference Manual. Bethesda (MD): NLM; 2009. Available from: http://www.ncbi.nlm.nih.gov/books/NBK9676/
16. McCray AT, Nelson SJ. The representation of meaning in the UMLS. Methods Inf Med. 1995; 34:193–201. [PubMed: 9082131]
17. McCray, AT.; Bodenreider, O. A conceptual framework for the biomedical domain. In: Green, R.; Bean, CA.; Myaeng, SH., editors. The Semantics of Relationships: An Interdisciplinary Perspective. Boston: Kluwer Academic Publishers; 2002. p. 181-198.
18. Hardiker NR, Coenen A. Interpretation of an international terminology standard in the development of a logic-based compositional terminology. Int J Med Inform. 2007; 76 Suppl 2:S274–S280. [PubMed: 17600764]
19. International Organization for Standardization (ISO). Health Informatics – Integration of a reference terminology model for nursing (ISO/TX 18104:2003). Geneva, Switzerland: ISO; 2003.
20. Matney SA, DaDamio R, Couderc C, Dlugos M, Evans J, Gianonne G, Haskell R, Hardiker N, Coenen A, Saba VK. Translation and integration of CCC nursing diagnoses into ICNP. J Am Med Inform Assoc. 2008; 15:791–793. [PubMed: 18755991]
21. Kim TY, Coenen A. Toward harmonising WHO International Classifications: a nursing perspective. Inform Health Soc Care. 2011; 36:35–49. [PubMed: 21208092]
22. Bodenreider, O.; Bean, CA. Relationships among knowledge structures: Vocabulary integration within a subject domain. In: Bean, CA.; Green, R., editors. Relationships in the Organization of Knowledge. Boston: Kluwer Academic Publishers; 2001. p. 81-98.
23. Grabar N, Varoutas PC, Rizand P, Livartowski A, Hamon T. Automatic acquisition of synonym resources and assessment of their impact on the enhanced search in EHRs. Methods Inf Med. 2009; 48:149–154. [PubMed: 19283312]
24. Meystre S, Haug PJ. Natural language processing to extract medical problems from electronic clinical documents: performance evaluation. J Biomed Inform. 2006; 39:589–599. [PubMed: 16359928]
25. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. J Am Med Inform Assoc. 2004; 11:392–402. [PubMed: 15187068]
26. Huang Y, Lowe HJ, Klein D, Cucina RJ. Improved identification of noun phrases in clinical radiology reports using a high-performance statistical natural language parser augmented with the UMLS specialist lexicon. J Am Med Inform Assoc. 2005; 12:275–285. [PubMed: 15684131]
27. Cimino JJ, Min H, Perl Y. Consistency across the hierarchies of the UMLS Semantic Network and Metathesaurus. J Biomed Inform. 2003; 36:450–461. [PubMed: 14759818]
28. Hardiker NR, Casey A, Coenen A, Konicek D. Mutual enhancement of diverse terminologies. AMIA Annu Symp Proc. 2006:319–323. [PubMed: 17238355]
29. College of American Pathologists (CAP), SNOMED CT® Mappings to NANDA, NIC, and NOC Now Licensed for Free Access Through National Library of Medicine. 2005. Available from: http://www.cap.org/apps/cap.portal?_nfpb=true&cntvwrPtlt_actionOverride=%2Fportlets%2FcontentViewer%2Fshow&_windowLabel=cntvwrPtlt&cntvwrPtlt%7BactionForm.contentReference%7D=media_resources%2Fnewsrel_nlm_mappings.html&_state=maximized &_pageLabel=cntvwr
30. CliniClue®. The Clinical Information Consultancy Ltd. 2011. Available from: http://www.cliniclue.com/
31. McCray AT, Burgun A, Bodenreider O. Aggregating UMLS semantic types for reducing conceptual complexity. Stud Health Technol Inform. 2001; 84:216–220. [PubMed: 11604736]
32. Oxford dictionaries. UK: Oxford University Press; 2011. Available from: http://oxforddictionaries.com/

33. Cimino JJ. Auditing the Unified Medical Language System with semantic methods. J Am Med Inform Assoc. 1998; 5:41–51. [PubMed: 9452984]

34. Bodenreider O, McCray AT. Exploring semantic groups through visual approaches. J Biomed Inform. 2003; 36:414–432. [PubMed: 14759816]

35. Gu H, Perl Y, Geller J, Halper M, Liu LM, Cimino JJ. Representing the UMLS as an object-oriented database: modeling issues and advantages. J Am Med Inform Assoc. 2000; 7:66–80. [PubMed: 10641964]

36. Perl Y, Chen Z, Halper M, Geller J, Zhang L, Peng Y. The cohesive metaschema: a higher-level abstraction of the UMLS Semantic Network. J Biomed Inform. 2002; 35:194–212. [PubMed: 12669983]

37. Zhang L, Perl Y, Halper M, Geller J, Hripcsak G. A lexical metaschema for the UMLS semantic network. Artif Intell Med. 2005; 33:41–59. [PubMed: 15617981]

38. Chen Y, Gu H, Perl Y, Halper M, Xu J. Expanding the extent of a UMLS semantic type via group neighborhood auditing. J Am Med Inform Assoc. 2009; 16:746–757. [PubMed: 19567802]

39. National Library of Medicine. Nursing Problem List Subset of SNOMED CT. 2011. Available from: http://www.nlm.nih.gov/research/umls/Snomed/nursing_problemlist_subset.html

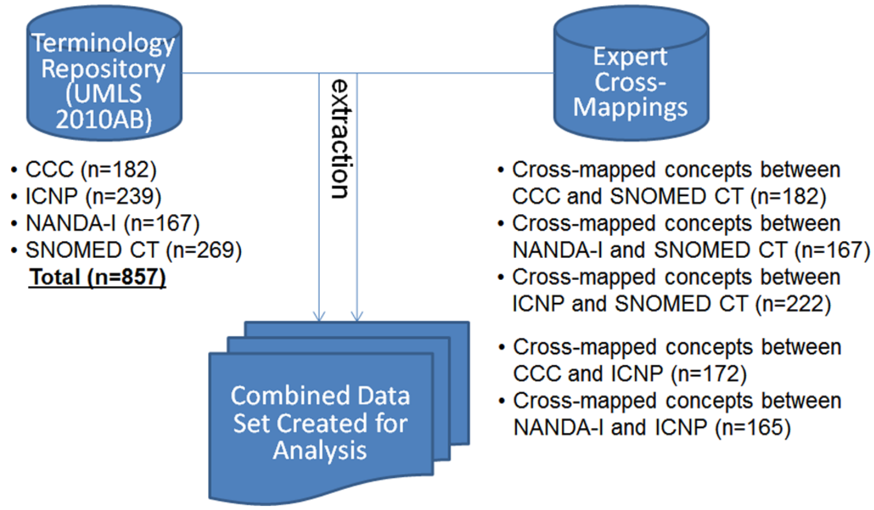

Figure 1.
Summary of Study Process

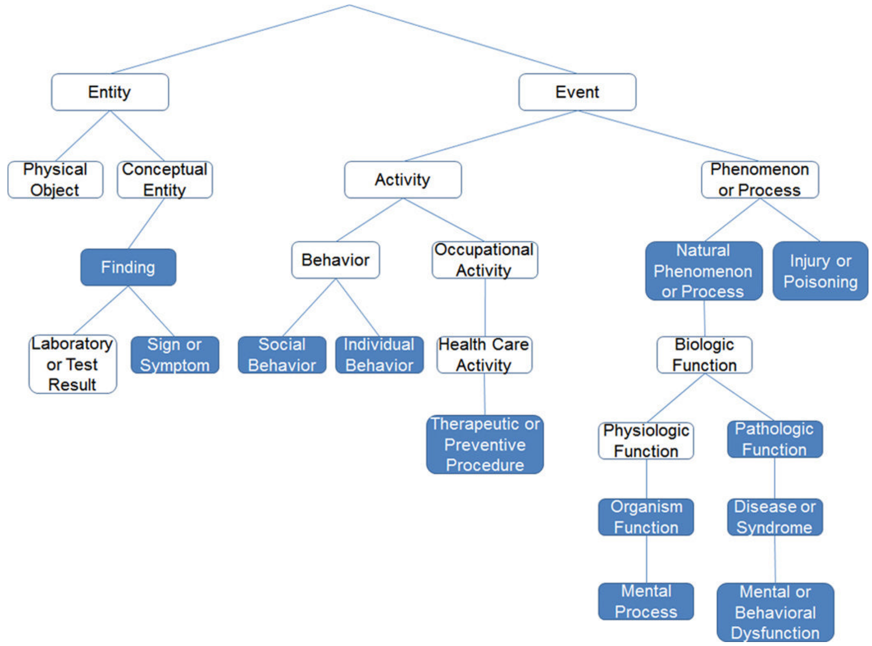

Figure 2.

Semantic Location of Nursing Diagnostic Concepts on the UMLS Semantic Network (Note that the figure presents a limited view of the semantic network to which the nursing diagnostic concepts belong. The ten boxes shaded include the nursing concepts examined in this study.)

Table 1

Nursing Diagnostic Concepts with Duplicate CUIs in UMLS (2010AB)

Source Terminology	CUI	Source Code	Source Concept
CCC	C0150027	E11	Family Coping Impairment
	C0150027	E11.1	Compromised Family Coping
	C0582456	N33	Injury Risk
	C0582456	N33.5	Trauma Risk
ICNP	C0009806	10001392	Perceived Constipation
	C0009806	10000567	Constipation
	C0582456	10015360	Risk for Injury
	C0582456	10015146	Risk for Trauma
NANDA-I	C0582456	00035	Risk for Injury
	C0582456	00038	Risk for Trauma

Table 2

Semantic Mappings of Nursing Diagnostic Concepts to SNOMED CT

Source Terminologies	No. of Concepts Examined	Cross-Mapped by Human Experts (%)	Cross-Mapped by CUIs in UMLS (%)	Concordance of UMLS Mapping
CCC and SNOMED CT	182	182 (100.0)	162 (89.0)	89%
NANDA-I and SNOMED CT	167	167 (100.0)	163 (97.6)	98%
ICNP and SNOMED CT	239	222 (92.8)	66 (27.6)	30%

Table 3

Semantic Mappings of Nursing Diagnostic Concepts to ICNP

Source Terminologies	No. of Concepts Examined	Cross-Mapped by Human Experts (%)	Cross-Mapped by CUIs in UMLS (%)	Concordance of UMLS Mapping
CCC and ICNP	182	172 (94.5)	41 (22.5)	24%
NANDA-I and ICNP	167	165 (98.8)	66 (39.5)	40%

Table 4

Examples of Matched and Non-matched Nursing Diagnostic Concepts in UMLS

CUI	Source Code	Source Concept	CUI	ICNP Code	ICNP Concept	Expert	UMLS CUI
C0150034	L26.3	Gas Exchange Impairment	C0150034	10001177	Impaired Gas Exchange	Exact match	Exact Match
C0027497	00134	Nausea	C0027497	10000859	Nausea	Exact match	Exact Match
C0015732	00014	Bowel Incontinence	C2712028	10027718	Bowel Incontinence	Exact match	No Match
C0184566	Q45	Comfort Alteration	C2364135	10023066	Discomfort	Exact match	No Match
C0231369	O38	Self Care Deficit	C2364166	10023410	Self Care Deficit	Exact match	No Match
C0150083	K25.4	Thermoregulation Impairment	C1998659	10001337	Impaired Thermoregulation	Exact match	No Match

Table 5

Synonymous Relationships of Nursing Diagnostic Concepts in UMLS (2010AB)

Source Terminology	Number of Concepts with a Synonymous Relationship (%)				Number of Concepts without a Synonymous Relationship (%)
	CCC	NANDA	ICNP	SNOMED CT	
CCC (n=182)	0	0	66 (36.3)	39 (21.4)	100 (54.9)
NANDA-I (n=167)	0	0	81 (48.5)	33 (19.8)	79 (47.3)
ICNP (n=239)	0	0	75 (31.4)	19 (7.9)	220 (92.1)

Table 6

Distributions of Nursing Diagnostic Concepts on the UMLS Semantic Network and Group

	Semantic Type	CCC (%) n=182	ICNP (%) n=239	NANDA (%) n=167	Semantic Group [31]
Entity	Finding	121 (66.5)	195 (81.6)	118 (70.7)	Disorders
	Sign or Symptom	15 (8.2)	8 (3.3)	10 (6.0)	Disorders
Event	Social Behavior	1 (0.5)	1 (0.4)	1 (0.6)	Activities & Behaviors
	Individual Behavior	-	1 (0.4)	1 (0.6)	Activities & Behaviors
	Therapeutic or Preventive Procedure	1 (0.5)	-	-	Procedures
	Natural Phenomenon or Process	-	1 (0.4)	-	Phenomena
	Organism Function	1 (0.5)	-	1 (0.6)	Physiology
	Mental Process	4 (2.2)	9 (3.7)	3 (1.8)	Physiology
	Pathologic Function	9 (4.9)	6 (2.5)	8 (4.8)	Disorders
	Disease or Syndrome	13 (7.1)	3 (1.3)	10 (6.0)	Disorders
	Mental or Behavioral Dysfunction	16 (8.8)	14 (5.9)	14 (8.4)	Disorders
	Injury or Poisoning	1 (0.5)	1 (0.4)	1 (0.6)	Disorders